Journal of Physics: Conference Series

Segmentation of Relevant Region in Breast Histopathology Images using FCM with Guided Initialization

To cite this article: Tan Xiao Jian *et al* 2019 *J. Phys.: Conf. Ser.* **1372** 012028

Segmentation of Relevant Region in Breast Histopathology Images using FCM with Guided Initialization

Tan Xiao Jian[1], Mustafa Nazahah[1*], Mashor Mohd Yusoff[1], Ab Rahman Khairul Shakir[2]

[1]University Malaysia Perlis, 02600 Arau, Perlis, Malaysia.
[2]Hospital Tuanku Fauziah, Jalan Tun Abdul Razak, 01000 Kangar, Perlis, Malaysia.

*nazahah@unimap.edu.my

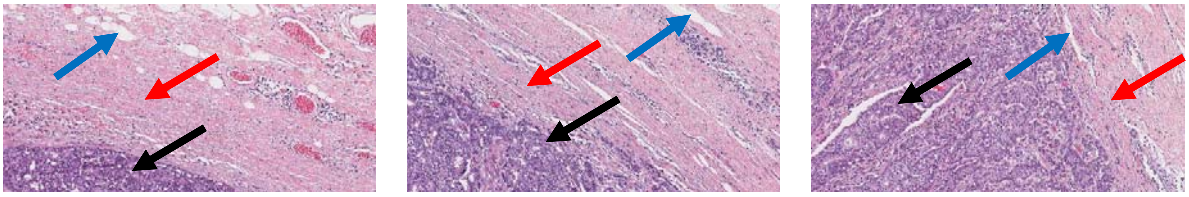
Abstract. This study proposes a modified initialization approach for the conventional FCM, namely FCM with guided initialization. The FCM with guided initialization was implemented to segment the relevant regions in the breast histopathology images. The initialization method to select initial centers is based on the Cyan (C) channel histogram. Area Overlap Measure (AOM) and Combined Equal Importance (CEI) were used to evaluate the performance of the proposed FCM with guided initialization. The obtained AOM and CEI for the overall dataset achieved promising results: 0.89 in AOM and 0.88 in CEI. When comparing the number of iterations required to complete the proposed FCM clustering algorithm, the FCM with guided initialization is found to be effective in reducing the search space by showing a lower number of iterations.

1. Introduction

Segmentation is often used as a preliminary step in the image processing, specifically in the breast histopathology image. The intrinsic nature of the breast histopathology image such as fuzziness, complexity, and heterogeneous [1, 2] elevate challenges in the segmentation. Segmentation of relevant regions (i.e., tumor regions) is crucial for breast carcinoma grading specifically in the assessment of glandular formation [3, 4]. The relevant regions provide diagnostically important information which hold a paramount role in providing scores to the overall grade.

Clustering is a common technique employed in data mining and data analysis [5]. Clustering technique performs classification or segmentation on objects that share similar characteristics, features and properties into subsets of data known as clusters. Theoretically, clustering algorithms aim to maximize similarity between intra-cluster objects and minimized similarity between inter-cluster objects. Fuzzy C Mean (FCM) is one of the popular approaches in clustering analysis. FCM has been reported by many studies on its superior ability in partition fuzzy dataset into different clusters with similar properties [5, 6]. However, FCM is not flawless. This algorithm is sensitive to initial guess. Dead center, center redundancy and trap in local minima are some typical limitations of FCM [7].

The main objective of this study is to reduce limitations of the conventional FCM algorithm. The conventional approach to obtain the initial centers for each cluster was modified so that the initial centers can be selected systematically from the intensity histogram of the input images. The proposed method for center initialization with FCM algorithm was implemented on the breast histopathology images to obtain relevant regions, non-relevant regions, and background areas (Figure 1). The organization of this study is as follows: Section 2, methodology; Section 3, experimental results; and Section 4, conclusion.



(a) (b) (c)

Figure 1. Samples of histopathology images. The relevant regions, non-relevant regions and background are shown by the black, red and blue arrows, respectively.

2. Methodology

Figure 2 shows the block diagram of the proposed method. The proposed method consists of four steps: color normalization, selection of color channel, FCM with guided initialization and post processing. The details of each sub-step are described in the following subsections.

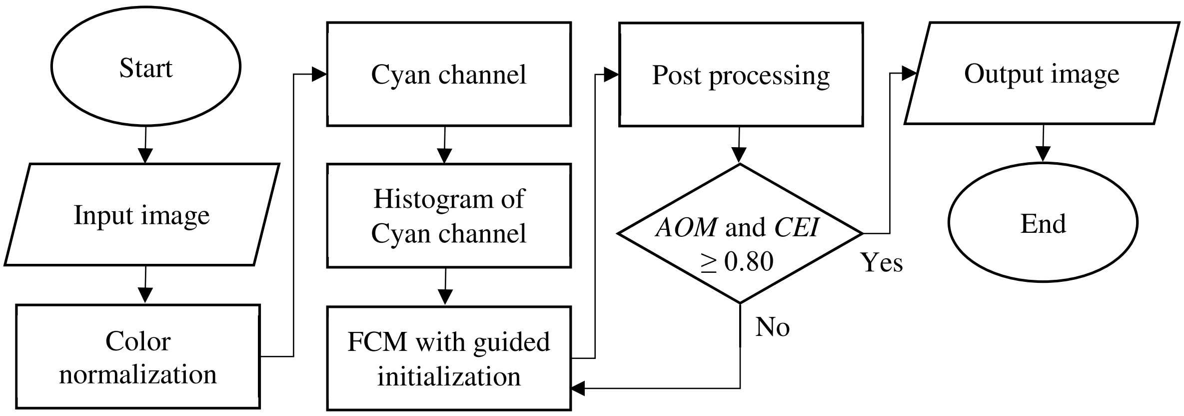


Figure 2. Block diagram of the proposed method.

2.1 Color Normalization

Color normalization step aims to standardize and reduce color inconsistency in the RGB input histopathology images. This step was done using histogram matching algorithm [8]. Histogram matching method was selected as it can provide fast and simple computation on large size of image and it can ensure smoothness of the clustering process in the later sub-step of segmentation.

2.2 Selection of color channel

For the purpose of relevant region segmentation, a comprehensive empirical study was conducted by investigating YCbCr, HSV, and CMY color models. The empirical analysis found that the Cyan (C) channel from the CMY color model shows significant discrimination in color features between the relevant, non-relevant regions, and the background areas. The C channel which complements the R channel (i.e., based on the RGB to CMY color conversion) provides a relatively high C value for the relevant regions, moderate C value for irrelevant regions, and a low C value for the background areas. Hence, the C channel was opted as the input color channel to the consequence sub-steps.

2.3 FCM with Guided Initialization

Based on the domain knowledge, the relevant and non-relevant regions in the breast histopathology image have a distinct degree of stain absorption, therefore, the relevant and non-relevant regions could be discriminated using the color feature [9]. The FCM clustering algorithm was selected to partition the input image into relevant regions, non-relevant regions, and background. The conventional FCM clustering algorithm applied with random initialization method have been reported with several limitations as described in Section 1. To address these limitations, the initialization method was modified to select initial centers based on the C channel histogram. Incorporation of the domain knowledge in color feature, specifically in the selection of initial centers for the relevant regions, non-relevant regions, and background are hypothesized to reduce the search space and avoid the limitations aforementioned. Three initial centers were selected for background (*IC1*), non-relevant regions (*IC2*), and relevant regions (*IC3*). The search for IC1 and IC3 were performed using the local maxima search algorithm (i.e., hill climbing method [10]) to obtain the first and second intensity peaks (Figure 3) whereas *IC2* was obtained using Equation (1).

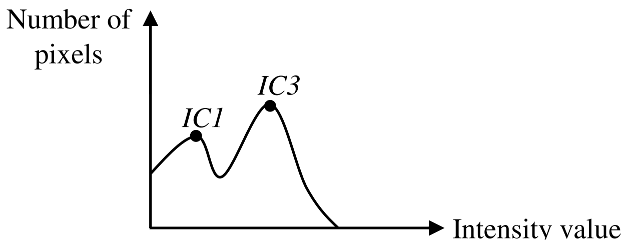


Figure 3. Illustration of C channel histogram.

$$IC2 = IC1 + \frac{IC3 - IC1}{2} \tag{1}$$

FCM clustering algorithm requires two input parameters: number of clusters (n_c) (i.e., n_c=3 in this study) and initial cluster value for each cluster. Consider an image with N number of pixels to be clustered into n_c regions. V_i is i^{th} data and C_j is j^{th} center (i = 1, 2, …, N and j = 1, 2, …, n_c). The steps of the FCM are as follows:

1. Set number of cluster (n_c).
2. Set initial cluster value for each cluster obtained using the proposed initialization method.
3. Calculate the membership (M_{ji}) using Equation (2):

$$M_{ji} = \frac{1}{\sum_{k=1}^{n_c}\left(\frac{d_{ji}}{d_{ki}}\right)^2}; if\ d_{ki} > 0, \forall j, i \tag{2}$$

where $d_{ji} = \left\|V_i - C_j\right\|$

$$\left.\begin{array}{l} M_{ji} = 1 \\ M_{ji} = 0; for\ i \neq k \end{array}\right\} if\ d_{ki} = 0$$

4. Calculate the new center using Equation (3):

$$C_j = \frac{\sum_{i=1}^{N} M_{ji}^2 V_i}{\sum_{i=1}^{N} M_{ji}^2} \quad (3)$$

5. Repeat steps 3 and 4 until the center remain unchanged.

2.4 Post Processing

The small artefacts on the FCM with guided initialization outputs were removed using morphological operation [11]. A "closing" morphological operation with a "disk" structure element (i.e., radius of 1 pixel) was implemented to remove and fill holes in the obtained outputs.

2.5 Evaluation Metrics

Two statistical metrics were used for evaluation: Area Overlap Measure (AOM) and Combined Equal Importance (CEI). AOM is defined as a ratio of the intersection to the union of the two areas to be compared. This metric is commonly used to evaluate the performance of the object region segmentation algorithm. CEI is an equation that combined AOM, over-segmentation, and under-segmentation to provide them an equal importance. The equations of AOM and CEI are given in Equations (4) and (5) [12], where A denotes to the result obtained from the proposed procedure and B denotes the ground truth.

$$AOM = \frac{area|A \cap B|}{area|A \cup B|} \quad (4)$$

$$CEI = \frac{AOM + \left(1 - \frac{area|B| - area|A \cap B|}{area|B|}\right) + \left(1 - \frac{area|A| - area|A \cap B|}{area|A|}\right)}{3} \quad (5)$$

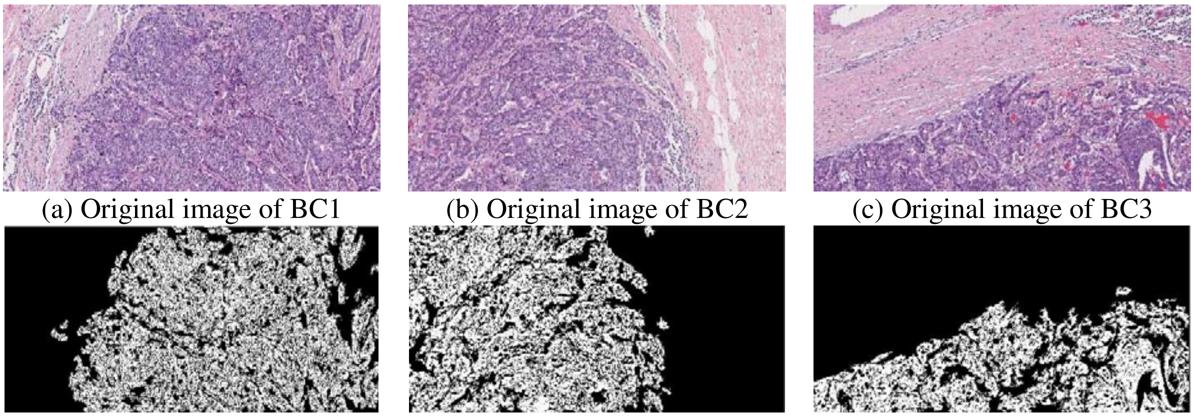
3. Experimental Results

3.1 Dataset

The histopathology slides of breast carcinoma were locally collected, in Malaysia. The breast histopathology slides were obtained from the Pathology Department, Hospital Tuanku Fauziah, Kangar, Perlis, Malaysia. A total of five breast histopathology slides were used in this study. These slides were prepared under a standard staining procedure aligned with the guideline of MOH, Malaysia, as in [13], by using the Hematoxylin and Eosin (H&E) stain. An Aperio CS2 WSI scanner was used to convert the slides into digital form (i.e., digital histopathology slides). The digital histopathology slides were viewed and captured using the Aperio ImageScope software at 10x magnification. The captured histopathology image is an 8-bit RGB colour with dimensions of 614x1240 pixels (size of pixel: 0.2521 µm per pixel) and is in tiff file format. A total of 25 histopathology images were used in this study, such that five images were captured from each slide which corresponding to different dominant regions.

3.2 Results and Discussion

In Figure 4, images (a), (b) and (c) show three RGB breast histopathology images namely, BC1, BC2 and BC3, respectively. Images (d), (e) and (f) show the outputs of the FCM with guided initialization applied on BC1, BC2 and BC3, respectively. The clustering output of *IC3* (i.e., relevant regions) are shown in white areas. The obtained AOM and CEI for BC1, BC2, BC3 and the overall dataset are given in Table 1. From the table, the proposed method is found to be robust and effective in segmenting the relevant regions. The obtained AOM and CEI in the overall dataset show promising results: 0.89 and 0.88, respectively.



(a) Original image of BC1 (b) Original image of BC2 (c) Original image of BC3

(d) Clustered image of BC1 (e) Clustered image of BC2 (f) Clustered image of BC3

Figure 4. Results of FCM clustering using the proposed initialization method.

Table 1. The obtained AOM and CEI for BC1, BC2, BC3 and the overall dataset.

Images	AOM	CEI
BC1	0.84	0.86
BC2	0.86	0.88
BC3	0.85	0.89
Average for 25 images	0.89	0.88

The number of iterations for the conventional FCM (with random initialization) and FCM with proposed guided initialization were compared and shown in Table 2. The FCM with proposed guided initialization shows its ability to reduce the search space and complete the clustering algorithm with a smaller number of iterations (i.e., shorter computation time). As the initial centers were selected systematically from the histogram of Cyan channel, the limitations such as dead center, center redundancy and local minima could be avoided.

Table 2. The number of iterations for BC1, BC2, BC3 and the overall dataset.

Images	Iteration(s)	
	Conventional FCM	FCM with guided initialization
BC1	12	7
BC2	12	5
BC3	11	6
Average for 25 images	13	6

4. Conclusion

This study presented a modified initialization method for the conventional FCM clustering algorithm, namely FCM with guided initialization. The initial centers of the FCM were not inherent from random generation, but selected systematically from the histogram of the Cyan channel. Domain knowledge in breast histopathology image was incorporated into the algorithm to assist the selection of initial centers from the histogram. The FCM with guided initialization is found to be robust and effective in the segmentation of relevant regions in breast histopathology images (Table 1). The proposed method is able to reduce the search space by completing the clustering algorithm with the lower number of iterations (Table 2). The proposed method is also able to avoid limitations such as dead center, center redundancy and trap in local minima as reported in Section 1. The similar initialization approach could be used in the other cluster analysis such as K-Mean clustering algorithm to address the similar limitations. It is important to emphasize that for different medical images (e.g., magnetic resonance (MR) images), the FCM with guided initialization might require different domain knowledge in selecting the initial center.

Acknowledgments
The authors would like to acknowledge the support from the Fundamental Research Grant Scheme (FRGS) under a grant number of FRGS/1/2016/SKK06/UNIMAP/02/3 from the Ministry of Higher Education Malaysia. The protocol of this study has been approved by the Medical Research and Committee of National Medical Research Register (NMRR) Malaysia (NMRR-17-281-34236).

References
[1] Ramírez-torres A, Rodríguez-ramos R, Sabina FJ, García-reimbert C, Penta R, Merodio J and et al 2017 The Role of Malignant Tissue on the Thermal Distribution of Cancerous Breast. *J Theor Biol*;426:152–61. doi:10.1016/j.jtbi.2017.05.031.
[2] Samuel SM, Varghese E, Varghese S and Büsselberg D 2018 Challenges and Perspectives in the Treatment of Diabetes Associated Breast Cancer. *Cancer Treat Rev*;70:98–111. doi:10.1016/j.ctrv.2018.08.004.
[3] Jian TX, Mustafa N, Mashor MY and Shakir K 2018 Hyperchromatic Nucleus Segmentation on Breast Histopathological Images for Mitosis Detection. *J Telecommun Electron Comput Eng*;10:27–30.
[4] Jian TX, Mustafa N, Mashor MY and Shakir K 2018 Segmentation Based Classification for Mitotic Cells Detection on Breast Histopathological Images. *J Telecommun Electron Comput Eng*;10:2–5.
[5] Cebeci Z and Yildiz F 2015 Comparison of K-Means and Fuzzy C-Means Algorithms on Different Cluster Structures. *J Agric Informatics*;6:13–23. doi:10.17700/jai.2015.6.3.196.
[6] Velmurugan T 2012 Performance Comparison between k-Means and Fuzzy C-Means Algorithms using Arbitrary Data Points. *Wulfenia J*;19:234–41.
[7] Jian TX, Mustafa N, Mashor MY and Shakir K 2018 An Improved Initialization Based Histogram of K-Mean Clustering Algorithm for Hyperchromatic Nucleus Segmentation in Breast Carcinoma Histopathological Images. *Lect Notes Electr Eng* 2018:1–6.
[8] Arai K, Kadoya N, Kato T, Endo H, Komori S, Abe Y and et al 2017 Feasibility of CBCT-based proton dose calculation using a histogram-matching algorithm in proton beam therapy. *Phys Medica*;33:68–76. doi:10.1016/j.ejmp.2016.12.006.
[9] Tan XJ, Mustafa N, Rahman KSA, Mashor MY and Ang WC 2018 Simple Landscapes Analysis for Relevant Regions Detection in Breast Carcinoma Histopathological Images. *Int. Conf. Comput. Approach Smart Syst. Des. Appl., IEEE*; p. 1–5. doi:10.1109/ICASSDA.2018.8477610.
[10] Ganesan P, Kalist V and Sathish BS 2016 Histogram based Hill Climbing Optimization for the Segmentation of Region of Interest in Satellite Images. *IEEE WCTFTR - Proc. World Conf. Futur. Trends Res. Innov. Soc. Welf.*, p. 5–9. doi:10.1109/STARTUP.2016.7583961.
[11] Mohammad S, Hasan A and Ko K 2016 Depth edge detection by image-based smoothing and morphological operations. *J Comput Des Eng*;3:191–7. doi:10.1016/j.jcde.2016.02.002.
[12] Abbas Q, Celebi ME and Garcia IF 2013 Breast mass segmentation using region-based and edge-based methods in a 4-stage multiscale system. *Biomed Signal Process Control*;8:204–14. doi:10.1016/j.bspc.2012.08.003.
[13] Guideline CP, editor 2010 Management of Breast Cancer. 2nd ed. *Ministry of Health Malaysia*.